PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G10L 7/02

A1

(11) International Publication Number: WO 98/39768

(43) International Publication Date: 11 September 1998 (11.09.98)

(21) International Application Number: PCT/SE98/00280

(22) International Filing Date: 17 February 1998 (17.02.98)

(30) Priority Data:
9700772-8 3 March 1997 (03.03.97) SE

(71) Applicant: TELEFONAKTIEBOLAGET LM ERICSSON (publ) [SE/SE]; S-126 25 Stockholm (SE).

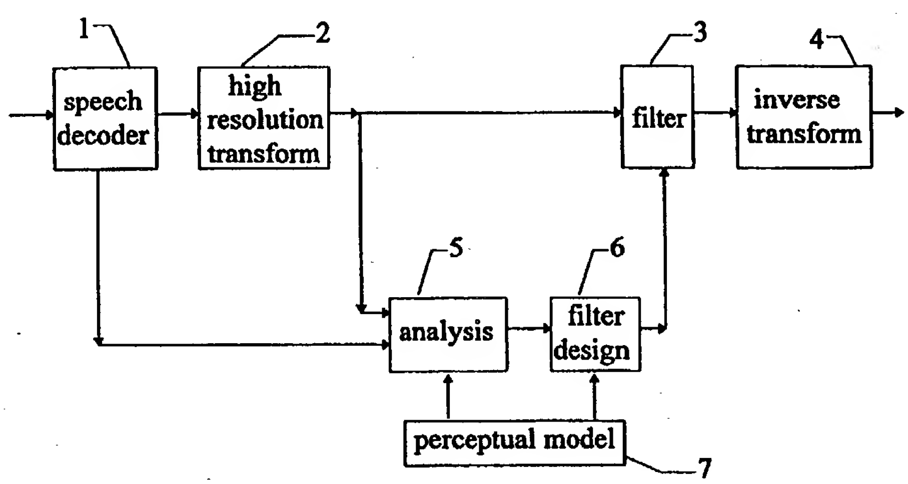
(72) Inventors: EKUDDEN, Erik; Fjärilsvägen 23, S-184 38 Åkersberga (SE). HAGEN, Roar; Kungsklippan 12, S-112 25 Stockholm (SE). KLEIJN, Bastiaan; Önnemovägen 59, S-146 53 Tullinge (SE).

(74) Agent: ERICSSON RADIO SYSTEMS AB; Common Patent Dept., S-164 80 Stockholm (SE).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: A HIGH RESOLUTION POST PROCESSING METHOD FOR A SPEECH DECODER



(57) Abstract

A post-processing method for a speech decoder (1) which gives a decoded speech signal in the time domain in order to obtain high frequency resolution from a frequency spectrum having non-harmonic and noise deficiencies. The method comprises the following steps: a) transforming (21) the decoded time domain signal to a frequency domain signal by means of a high frequency resolution transform (FFT); b) analysing (5) the energy distribution of said frequency domain signal throughout its frequency area (4 kHz) to find the disturbing frequency components and to prioritize such frequency components which are situated in the higher part of the frequency spectrum; c) finding (6) the suppression degree for said disturbing frequency components based on said prioritizing; d) controlling a post-filtering (31) of said transform in dependence of said finding (6); and e) inverse transforming (4) the post-filtered transform in order to obtain a post-filtered decoded speech signal in the time domain.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

A HIGH RESOLUTION POST PROCESSING METHOD FOR A SPEECH DECODER.

TECHNICAL AREA

The present invention relates to a post processing method for a speech decoder to obtain a high frequency resolution. The speech decoder is preferably used in a radio receiver for a mobile radio system.

DESCRIPTION OF PRIOR ART

In speech and audio coding it is common to employ post-processing techniques in the decoder in order to enhance the perceived quality of the decoded speech.

Post-processing techniques, such as traditional adaptive postfiltering, are designed to provide perceptual enhancements by emphasising formant and harmonic structures and to some extent de-emphasise formant valleys.

The present invention proposes a novel technique for post-processing which includes a high resolution analysis stage in the decoder. The new technique is more general in terms of noise reduction and speech enhancements for a wide range of signals including speech and music.

There is no known solution to a post-processing scheme for speech or audio coders which uses an analysis of the received parameters and the spectrum of the received signal to estimate a more precise coding noise level, combined with highly (non-harmonic) frequency selective de-emphasis filtering.

The formant postfilters in LPC based coders where the filter is derived from the received LPC parameters are well known. It does not make use of the spectral fine structure, and provides very limited frequency resolution.

Various types of LTP postfilters are well known. These filters can only affect the overall harmonic structure of the decoded signal, and can although providing high frequency resolution not address non-harmonic localised coding noise or artifacts. They are also particularly tailored to speech signals.

It is also known that analysis of the decoded speech at the receiver side can be used to estimate parameters in for example a pitch postfilter. This is performed in the LD-CELP for example. This is however only a harmonic pitch postfilter, where the "analysis" is only aimed at finding the pitch harmonics. No overall analysis of where the actual coding noise problems and artifacts are located is performed.

Relatively frequency selective "postfilters" have also been proposed in the context of removing frequency regions not coded by a very low bit-rate coder [1].

SUMMARY OF THE INVENTION

Many speech coders, e.g. LPC-based analysis-by-synthesis (LPAS) coders, make use of an error criterion in the parameter search which has very limited frequency selectivity. Further, the waveform matching criterion in many such coders will limit the performance for low energy regions, such as the spectral valleys, i.e. the control of the noise distribution in these frequency areas is much less precise.

When spectral noise weighting is used in the coder, the overall error spectrum, i.e. the coding noise, is spectrally shaped, although limited by the frequency resolution of the weighting filter. However, there may still be spectral regions, typically in spectral valleys or other low energy regions, with relatively high noise or audible artifacts which limit the perceived quality. For a given bit-rate, coder structure and input signal, the coder can only achieve a certain noise level. The relatively poor frequency selectivity in the coder and the post-processing, and the limiting bit-rate can not attack the quality problem areas for all types of signals.

A traditional bandwidth expanded LPC formant postfilter with low order (typically 10^{th} order) has relatively low frequency selectivity and can not address localised noise or artifacts.

Harmonic pitch postfilters can provide high frequency resolution, but can only perform harmonic filtering, i.e. not localised non-harmonic filtering.

Speech and music signals, for example, have fundamentally different structures and should employ different post-processing strategies. This can not be achieved unless the received signal is analysed and high resolution selective filters are used in the post-processing. This is not done presently.

The object of the present invention is to obtain a high frequency resolution post-processing method for the decoded signal from a speech or audio decoding device which at least reduces not desired influence of the non-harmonics and other coding noise in the decoded frequency spectrum.

The decoded signal is analysed to find likely frequency areas with coding noise. The high-resolution analysis is performed on the spectrum of the decoded speech signal and based on knowledge about the properties of the speech coding algorithm combined with parameters from the speech decoder. The output of the analysis is a filtering strategy in terms of frequency areas where the signal is de-emphasised to reduce coding noise and enhance the overall perceived quality of the coded speech.

The method of the invention utilises a transform that gives a high frequency resolution spectrum description. This may be realized using the Fourier transform, or any other transform with a strong correlation to spectral content. The length of the transform may be synchronized with the frame length of the decoder (e.g. to minimise delay), but must allow for a sufficiently high frequency resolution.

After the transformation, analysis of the spectral content and decoder attributes is made in order to identify problem areas where the coding method introduced audible noise or

artifacts. The analysis also exploits a perceptual model of human hearing. The information from the decoder and the knowledge about the coding algorithm help estimate the amount of coding noise and its distribution.

The information derived in the analysis step and the perceptual model are used for a filter design in two steps:

The frequency areas to de-emphasise are determined.

The amount of filtering in each area is determined.

This gives a candidate filter which may be further refined in terms of dynamic properties. For instance, the filter characteristic may be unsuitable because it produces artifacts when used following previous filters. Also, the dynamic properties of the decoded signal can be taken into account by limiting the amount of change in the filtering as compared to how much the decoded signal is changing.

The strategy for filter design described above allows for very frequency selective postfiltering which is targeted at adaptively suppressing problem areas. This is in contrast to current general-purpose postfiltering that is always applied without a specific analysis. Furthermore, the method allows for different filtering for different types of signals such as speech and music.

The filtering of the decoded signal must be performed with high frequency resolution. The filter can for instance be implemented in the frequency domain and finally followed by an inverse transform. However, any alternative implementation of the filtering process may be used.

In an alternative low-delay implementation of the proposed solution, the filtering may be performed using the result

from the analysis and filter design obtained in previous frames only. The delay incurred by the alternative implementation of the solution could then be kept very low.

BRIEF DESCRIPTION OF THE DRAWINGS

The method according to the present invention will be described in detail with reference to the accompanying drawings in which

Figure 1 shows a block diagram of the different functional blocks to perform the method according to one embodiment of the present invention;

Figure 2 shows a block diagram of another embodiment of the method according to the present invention;

Figure 3 shows a more detailed block diagram of the analysis and the filter design of Figures 1 and 2; and

Figure 4 shows a diagram which illustrates the frequency spectrum of a decoded signal and the principles of the post-processing according to the present invention.

DESCRIPTION OF PREFERRED EMBODIMENTS

The following description illustrates a working implementation of the invention described above. It is designed for use with a CELP (Code Exited Linear Predictive) coder. Such coders tend to generate noise in low energy areas of the spectrum and especially in valleys between peaks that have a complex non-harmonic relation as, for instance, music. The following points and Figure 3 illustrate the detailed implementation.

Figure 1 is a block diagram of the various functions performed by the present invention. A speech decoder 1, for instance in a radio receiver of a mobile telephone system decodes an incoming and demodulated radio signal in which parameters for the decoder 1 have been transmitted over a radio medium.

On the output of the decoder a decoded speech signal is obtained. The frequency spectrum of the decoded signal has a certain characteristics due to the transmission and to the decoding characteristics of the speech decoder 1.

The decoded signal in the time domain is converted by a Fast Fourier Transformation FFT designated by block 2 so that a frequency spectrum of the decoded signal is obtained. This frequency spectrum together with the frequency characteristics of the speech decoder are analysed, block 5, and the result of the analysis is supplied to a filter design unit 6. This design unit 6 gives an information signal to the post-filter 3. This filter performs a post-filtering of the frequency spectrum of the speech signal in order to eliminate or at least reduce the influence of the noise components in the decoded speech signal spectrum. The spectrum signal from the filter 3 which is free from disturbing frequency components or at least with strongly reduced disturbing components, is fed to a block 4 where the inverse transformation to that in block 2 is performed.

A perceptual model 7 can be added to the analysis and the filter design which influences the filtering (block 3) of the decoded speech signal spectrum as desired. This does not form any essential part of the present method and is therefore not described further.

In general terms, the spectral content of the decoded signal is analyzed in the following way in order to obtain measures that are used for identifying areas to de-emphasise.

The envelope of the magnitude spectrum is estimated in order to separate the overall spectral shape from the high resolution fine structure. The envelope may be estimated by a peak-picking process using a sliding window of sufficient width.

Smoothing of the magnitude spectrum may be performed to avoid ripple.

The resulting two vectors are used to identify sufficiently narrow spectral valleys of a certain depth. This gives candidate areas where filtering may be applied.

The spectrum may also be analyzed using a perceptual model to obtain a noise masking threshold.

The attributes from the decoder are analyzed in order to estimate a likely distribution and level of noise or artifacts introduced by the specific coder in use. The attributes are dependent on the coding algorithm but may include for instance: spectral shape, noise shaping, estimated error weighting filter, prediction gains - for instance in LPC and LTP, bit allocation, etc. These attributes characterize the behaviour of the coding algorithm and the performance for coding the specific signal at hand.

All, or parts of, the information about the coded signal derived is output from the analysis 5 and used for filter design 6.

In Figure 2, another embodiment of the post-processing method is shown. The difference from Figure 1 is that the analysis 5 and the filter design 6 is carried out in the frequency domain, while the post-filtering 8 of the decoded speech signal is carried out in the time domain. The output of the filter design unit 6 gives an information/control signal but now to the time domain filter 8 instead of the frequency domain filter 3 above.

Figure 3 shows a more detailed block diagram than Figures 1 and 2 for illustrating the inventive method.

The output of the speech decoder 1 in, for instance, a radio receiver is connected to a functional block 21 performing a 256 point Fast Fourier Transformation (FFT). A 256-point FFT is then performed every 128 samples using a Hanning window. Thus, every 128 samples a new block is processed. The log-magnitude of the FFT transform is computed along with the phase spectrum (which is not processed).

The analysis (block 5) consists of:

Estimating the envelope of the log-magnitude spectrum by computing each frequency point as the maximum of the log-magnitude spectrum within a sliding window of length 200 Hz in each direction. Peak-picking on the resulting vector is done by finding the frequency points where the log-magnitude spectrum equals the maximum value vector. Linear interpolation is performed between the peaks to get the envelope vector.

Smoothing the log-magnitude spectrum by taking the maximum within a sliding window of length 75 Hz in each direction.

Estimating the slope of the spectrum.

The filter design (block 6) consists of determining the areas where the smoothed log-spectrum curve is lower than the log-magnitude envelope curve by more than a specific value. These areas are suppressed if they correspond to more than one consecutive frequency point. Furthermore, if the valley is deeper than a certain high value, the suppression is widened to include the entire area between the peaks. The amount of spectral suppression in the log-domain at each frequency point to be suppressed is determined by the slope such that low energy areas get more suppression. The formula used is linear in the log-domain with no suppression for the last 1 kHz at the low end of the suppression (i.e. for a low-pass slope, the first 1 kHz is not suppressed and the other way around for an high-pass slope). This is done because of the character of the CELP coder which tends to generate more noise for low energy frequency areas.

The squared distance of the log-magnitude spectrum between the current and previous spectrum is computed along with the same measure for the suppression vectors. If the ratio of the values for the suppression vector and the spectrum itself is higher than a certain value (i.e. the suppression changes relatively too much compared to the signal spectrum), the suppression vector is smoothed by simply replacing it by the average of the current and previous suppression.

The filtering operation (block 31) is performed by simply subtracting the amount of suppression determined in the previous point from the log-magnitude spectrum of the decoded signal.

The inverse transform (block 4) is performed by first reconstructing the Fourier transform from the log-magnitude spectrum resulting from the filtering and the phase spectrum as passed directly from the transform. Note that an overlap and add procedure is employed to avoid artifacts because of discontinuities between the analysis frames.

The analysis block 5 of Figure 1 consists in this embodiment of an envelope detector 51, a smoothing filter 52 and a slope detector 53.

From the envelope detector the envelope signal e of the FFT-spectrum is obtained as shown in the diagram of Figure 4. The smoothing filter 52 gives a signal s_m representing the smoothed frequency characteristic from the FFT, block 21.

The filter design unit 6 consists in this embodiment of a comparator unit 61, a suppressor 62 and a unit 63 performing a dynamic processing.

The two signals e and s_m from the analysis block 5 are combined in the comparator unit 61. The difference between signals e and s_m is compared with a fix threshold T_h in the comparator 61 in order to determine a non-desired formant valley and the associated frequency interval. A signal s_1 is obtained which contains information about these.

The suppressing value forming unit 62 is controlled by a signal s_2 obtained from the slope unit 53 in the analyse block 5. Signal s_2 indicates the slope and in dependence on the slope value more or less suppression is performed on the frequency spectrum determined by signal s_1.

The dynamic unit 63 performs an adaption of the suppression from one frame to another so that sudden increase in

suppression indicated in the output signal from the suppression unit 62 do not happen.

The filter 3 of Figure 1 is in the embodiment according to Figure 3 a filter 31 (corresponding to filter 3 in Fig 1), called a subtractor in Figure 3, which performs a spectral subtraction. The signal value obtained from the dynamic unit 63 is the suppression value and is then subtracted from the frequency spectrum characteristic obtained from the FFT unit 21 within the frequency intervals determined by the signal s_1 as above. The result will be that the disturbing valleys in the frequency spectrum from the speech decoder 1 are reduced to a desired value before the final inverse transformation in block 4.

Depending on the slope s_l of the frequency spectrum characteristic different average values of the spectrum magnitudes are obtained. The slope gives high magnitude values in the beginning of the frequency spectrum where the speech decoder 1 is "strong" i.e. is capable of decoding correctly independent of possible noise components in the spectrum. For higher frequencies, where the slope implies lower magnitude values of the spectrum characteristic, it is more important to perform a good suppression of the valleys in the characteristic.

The frequency diagram of Figure 4 is intended to illustrate this. The smoothed frequency spectrum s_m and its envelope e are compared as mentioned above and the difference is compared with a fix threshold T_h. This gives in this example at least two different frequency areas f_1 and f_2 around the frequencies f_1 and f_2, respectively for which the valleys v_1 and v_2 are regarded as disturbing i.e. due to non-

harmonics/disturbing noise which the speech decoder cannot handle. Only these two frequency areas have been illustrated in Figure 4 although several other such areas are present both in the lower and in the higher part of the frequency spectrum.

The signal s_1 from the comparator 61 carries information about what frequency areas f_1, f_2, ... are to be suppressed and the signal s_2 from the slope detector 53 carries information about how great suppression is to be made. As mentioned above, if the detected frequency area is situated in the beginning of the spectrum as, for instance f_1, the suppression can be low while for area f_2 which is situated in the upper band, the suppression should be greater.

The dynamic unit 63 is adapting the suppression from one speech block to another. Preferably the incoming speech block (128 points) are treated with overlap so that when half a speech block has been processed in the blocks 5 and 6, the processing of a new subsequent speech block is started in the analyser block 5.

The dynamic unit 63 gives thus a signal which represents correction values to be subtracted from the spectrum characteristic which is done in the subtractor 31 corresponding to filter 3 in Fig 1. The improved frequency spectrum of the speech signal is thereafter inverse transformed in the inverse Fast Fourier Transformer 4 as above described with respect to the overlapping speech blocks.

The method can also be applied to a signal internal to the speech or audio decoder. The signal will then be processed by the method and thereafter further used by the decoder to

produce the decoded speech or audio signal. An example is the excitation signal in a LPC coder which can be processed by the proposed signal before the decoded speech is reconstructed by the linear prediction synthesis filter.

The fact that the method de-emphasises frequency areas in the decoded signal can be exploited during encoding such that the coding effort can be re-directed from the de-emphasised areas. For instance, the error weighting filter of an LPAS coder can be modified to lessen the weighting of the error in de-emphasised areas in order to accomplish this. Thus, the method can be used in conjunction with a modified encoder which takes the post-processing introduced by the method into account.

Merits of the Invention

Possibility to suppress coding noise and artifacts at localised frequency areas with high resolution. This is particularly useful for complex signals such as music. The method significantly enhances sound quality for complex signals while also enhancing the quality of pure speech although more marginally.

References

[1] D. Sen and W. H. Holmes, "PERCELP - Perceptually Enhanced Random Codebook Excited Linear Prediction", in Proc. IEEE Workshop Speech Coding, Ste. Adele, Que., Canada, pp. 101-102, 1993

CLAIMS

1. A post-processing method for a speech decoder (1) which gives a decoded speech signal in the time domain in order to obtain high frequency resolution from a frequency spectrum having non-harmonic and noise deficiencies, comprising the steps of:

a) performing (2) a high-frequency resolution transform on the decoded signal to obtain a frequency spectrum of the decoded speech signal,

b) analysing (5) said frequency spectrum in terms of estimating the likely coding noise characteristics in various frequency areas (f_1, f_2), and

c) performing high frequency resolution filtering of said frequency spectrum based on the analysing step in order to at least significantly reduce the frequency components in said frequency areas.

2. The method in Claim 1, where said analysis (5) uses the decoded high resolution signal spectrum.

3. The method in Claim 2, where said analysis (5) exploits decoder attributes.

4. The method in Claim 2, where said analysis (5) exploits properties of the coding algorithm.

5. The method in Claim 2, where said analysis (5) exploits a perceptual model (7).

6. The methods in Claim 1 to 5, where said filtering exploits dynamic properties of the filter.

7. The method in Claim 6, where said filtering exploits dynamic properties of the decoded signal.

8. A post-processing method for a speech decoder (1) which gives a decoded speech signal in the time domain in order to obtain high frequency resolution from a frequency spectrum having non-harmonic and noise deficiencies,

c h a r a c t e r i z e d in the steps of:

a) transforming (21) the decoded time domain signal to a frequency domain signal by means of a high frequency resolution transform (FFT),

b) analysing (5) the energy distribution of said frequency domain signal throughout its frequency area (4 kHz) to find the disturbing frequency components and to prioritize such frequency components which are situated in the higher part of the frequency spectrum,

c) finding (6) the suppression degree for said disturbing frequency components based on said prioritizing,

d) controlling a post-filtering (31) of said transform in dependence of said finding (6), and

e) inverse transforming (4) the post-filtered transform in order to obtain a post-filtered decoded speech signal in the time domain.

9. Method according to claim 8,

c h a r a c t e r i z e d in that

said analysing (5) includes

a) detecting (51) the envelope of a signal representing said frequency spectrum and forming a corresponding envelope signal (e),

b) estimating (53) the slope of said signal representing the frequency spectrum and forming a corresponding slope signal (s_1), and that

said filter design (6) includes

c) comparing said signal representing the frequency spectrum with said slope signal (s_1) in order to locate said disturbing frequency components (f_1, f_2),

d) forming a value representing the suppression degree for a specific frequency component based on the result of said comparing and said signal (s_1) corresponding to the slope, and repeating said forming for a number of such specific components, giving a number of values, said values being used as a control of said post-filtering of the frequency spectrum signal.

10. Method according to claim 9,

c h a r a c t e r i z e d in that said signal representing the frequency spectrum is a smoothed (53) signal from the signal obtained after said transforming (21).

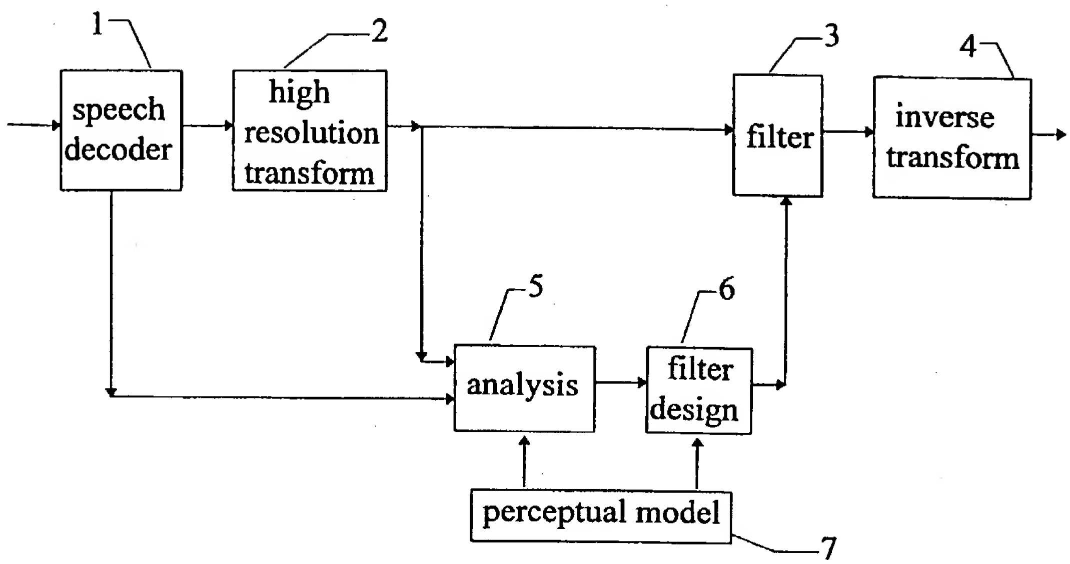

1
speech decoder
2
high resolution transform
3
filter
4
inverse transform
5
analysis
6
filter design
perceptual model
7

Fig.1

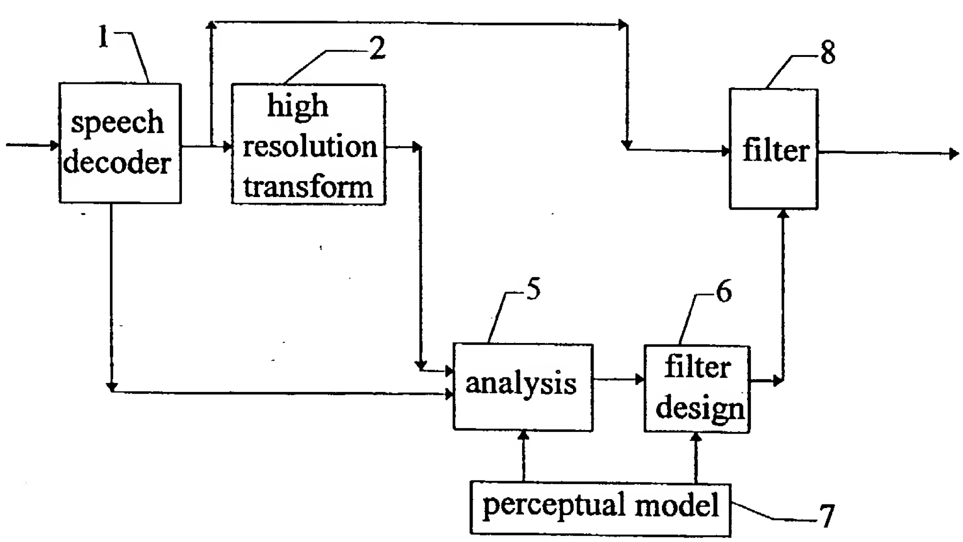

1
speech decoder
2
high resolution transform
8
filter
5
analysis
6
filter design
perceptual model
7

Fig.2

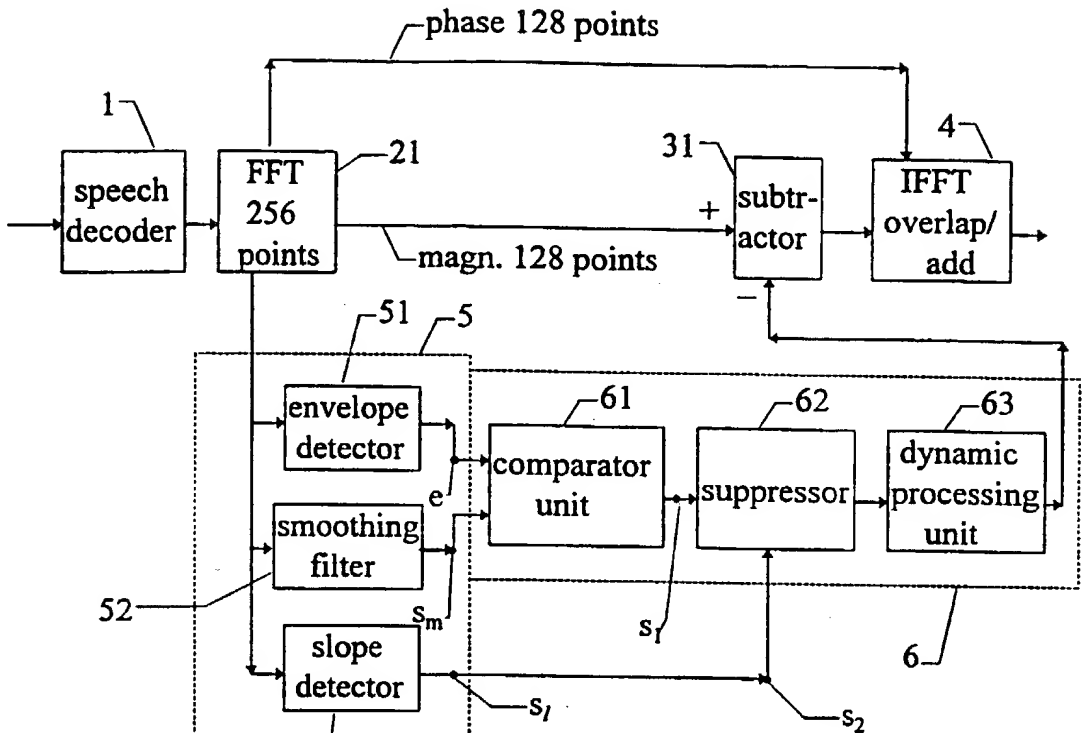

phase 128 points
1
speech decoder
FFT 256 points
21
magn. 128 points
31
+
subtr-actor
−
4
IFFT overlap/ add
51
5
envelope detector
e
61
comparator unit
62
suppressor
63
dynamic processing unit
52
smoothing filter
s_m
s_l
slope detector
s_l
s_2
6
53

Fig.3

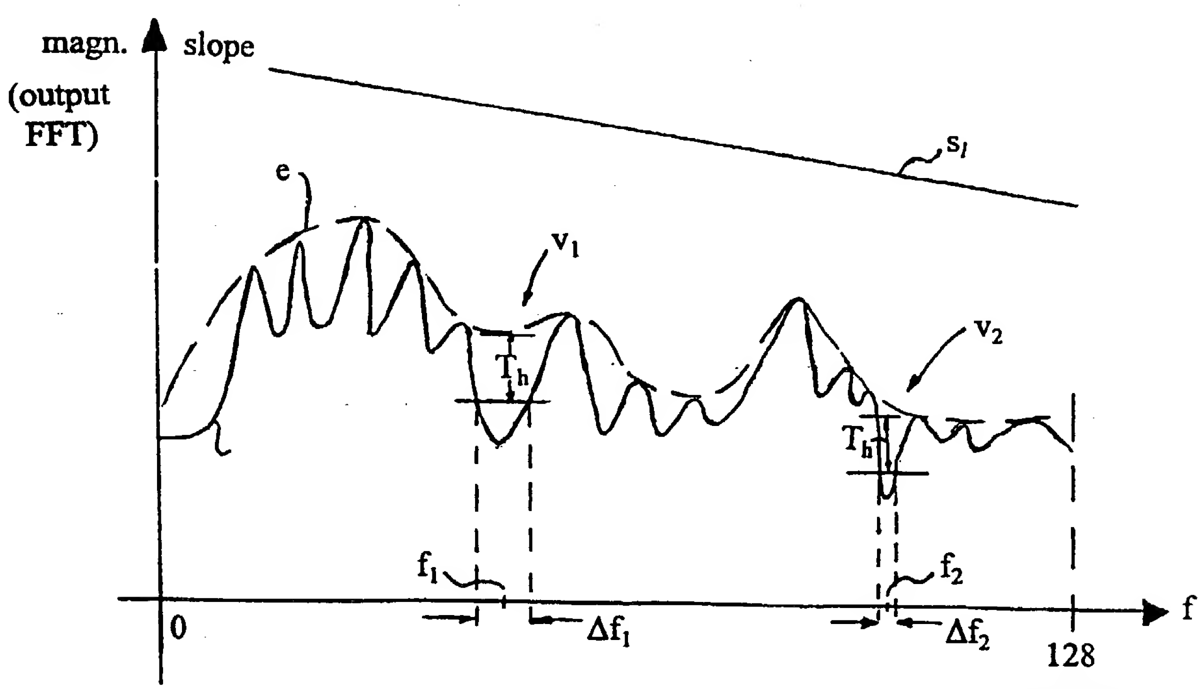

magn. (output FFT)
slope
s_l
e
v_1
v_2
T_h
T_h
f_1
f_2
0
Δf_1
Δf_2
128 (4 kHz)
f

Fig.4